

University of Groningen

Biochemical characterization and bioinformatic analysis of two large multi-domain enzymes from *Microbacterium aurum* B8.A involved in native starch degradation

Valk, Vincent

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Valk, V. (2017). *Biochemical characterization and bioinformatic analysis of two large multi-domain enzymes from *Microbacterium aurum* B8.A involved in native starch degradation*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

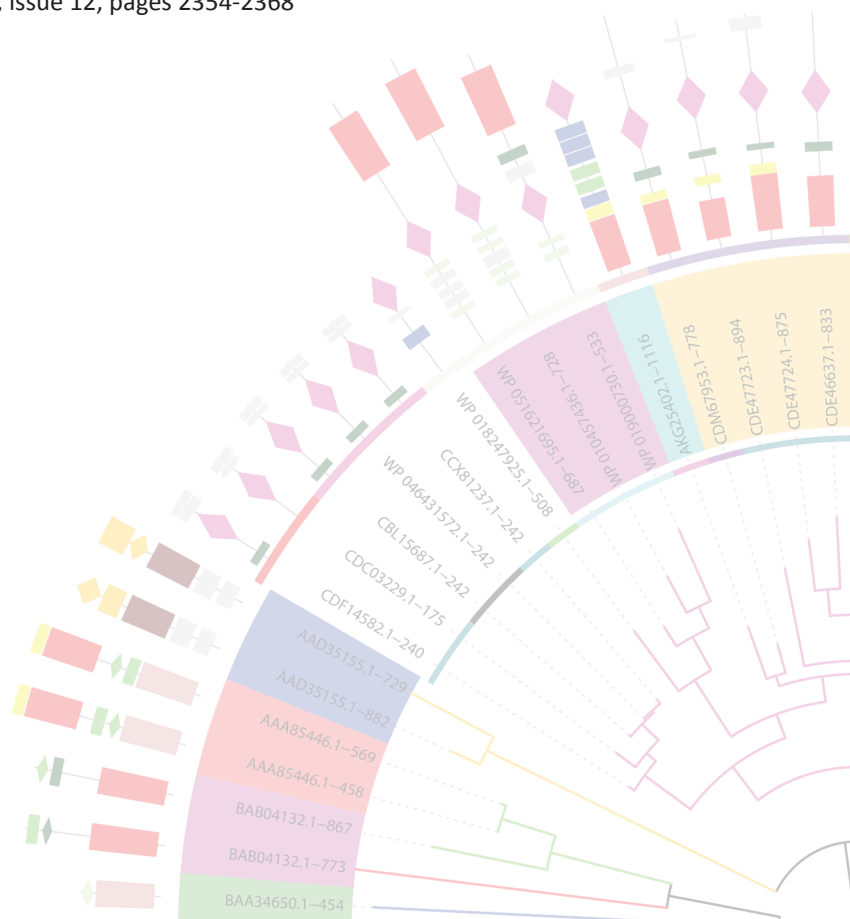
Carbohydrate Binding Module 74 is a novel starch binding domain associated with large and multi-domain α -amylase enzymes

Vincent Valk^{1,2}, Alicia Lammerts van Bueren¹, Rachel M. van der Kaaij¹, and Lubbert Dijkhuizen¹

¹Microbial Physiology Research Group, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Groningen, The Netherlands.

²Top Institute of Food and Nutrition (TIFN), Nieuwe Kanaal 9A, 6709 PA, Wageningen, The Netherlands.

This work has been published in The FEBS Journal (2016)
volume 283, issue 12, pages 2354-2368



Abstract

Microbacterium aurum B8.A is a bacterium that originates from a potato starch-processing plant and employs a GH13 α -amylase (MaAmyA) enzyme that forms pores in potato starch granules. MaAmyA is a large and multi-modular protein that contains a novel domain at its C-terminus (Domain 2). Deletion of Domain 2 from MaAmyA did not affect its ability to degrade starch granules but resulted in a strong reduction in granular pore size. Here, we separately expressed and purified this Domain 2 in *Escherichia coli* and determined its likely function in starch pore formation. Domain 2 independently binds amylose, amylopectin and granular starch but does not have any detectable catalytic (hydrolytic or oxidizing) activity on α -glucan substrates. Therefore we propose that this novel starch binding domain is a new carbohydrate binding module (CBM), the first representative of family CBM74, that assists MaAmyA in efficient pore formation in starch granules. Protein-sequence based BLAST searches revealed that CBM74 occurs widespread, but in bacteria only, and is often associated with large and multi-domain α -amylases containing family CBM25 or CBM26 domains. CBM74 may specifically function in binding to granular starches to enhance the capability of α -amylase enzymes to degrade resistant starches. Interestingly, the majority of family CBM74 representatives are found in α -amylases originating from human gut associated Bifidobacteria where they may assist in raw starch degradation. The CBM74 domain thus may have a strong impact on the efficiency of resistant starch digestion in the mammalian gastrointestinal tract.

Introduction

Starch is an abundantly available carbohydrate, present as storage material in plants [155]. It forms an important part of food and feed for humans and animals. Depending on the preparation of our food, native plant starch granules may still be present in the form of resistant starch; after heating mostly solubilized starch remains [12]. Also due to their high crystallinity, starch granules are relatively resistant against enzymatic degradation. Nevertheless, a diversity of bacteria is able to degrade granular starch, employing highly efficient α -amylase enzymes (www.cazy.org) [20]. α -Amylases acting on starch granules generally contain one or more Carbohydrate Binding Modules (CBMs). Such auxiliary domains may serve to position the enzyme active site into close and prolonged vicinity of starch granules, allowing hydrolysis of the insoluble substrate [47,156,157]. More recently, lytic polysaccharide monooxygenases (LPMOs) have been shown to oxidatively degrade insoluble polysaccharides, including cellulose, chitin and resistant starches. Starch LPMOs belong to family AA13 and thus far are only found in fungal species [158,159].

In previous work we have isolated the Gram-positive bacterium *Microbacterium aurum* strain B8.A from the sludge of a potato starch-processing factory on the basis of its ability to use granular starch as carbon- and energy source for growth. Extracellular enzymes hydrolyzing granular starch were detected in the growth medium of *M. aurum* B8.A [127]. Recently we reported the characterization of the raw starch degrading α -amylase MaAmyA enzyme of *M. aurum* B8.A (chapter 2, published as [160]). This very large α -amylase enzyme (1417 aa) carries multiple Carbohydrate Binding Modules (2 CBM25) and Fibronectin domains (4 FNIII) and initiates granular starch degradation by introducing pores (Fig. 1). At its C-terminus, MaAmyA carries a novel protein domain of 300 aa (Domain 2). A truncated MaAmyA variant in which Domain 2 was removed (MaAmyA7) remained fully active in starch granule degradation but introduced pores approximately 3 times smaller in size than full length MaAmyA (Fig. 1). Further deletions from the C-terminal end including the 2 CBM25 domains resulted in the loss of granular starch degradation ability (chapter 2, published as [160]).

Carbohydrate binding modules (CBMs) are non-catalytic protein modules associated with carbohydrate-active enzymes that bind to carbohydrate substrates and stimulate the catalytic efficiency of the enzyme [161]. CBMs are found in approximately 10% of all known glycoside hydrolase (GH) proteins recorded in the CAZy database [20,162], currently with a total of 71 CBM families. Starch binding domains (SBDs), constituting a CBM subgroup, are able to bind to starch. Currently SBDs have been found in CBM families 20, 21, 25, 26, 34, 41, 45, 48, 53, 58, 68 and 69 [163]. SBDs are usually 100-130 aa long [48,57,163] and mainly present in GH13 α -amylases, GH15 glucoamylases, GH77 amylomaltases and GH14 β -amylases [20,162]. The best studied SBDs are found in the CBM20

family. Next to binding, some SBD also play a role in the disruption of the starch structure [157] thereby making the polymer better accessible for hydrolysis by the catalytic domain [164,165]. Some α -amylases contain multiple binding domains, from the same or different CBM families, which generally results in an increase in starch binding capability of the enzyme [49,64].

In this study we expressed and purified Domain 2 from the MaAmyA enzyme in *Escherichia coli* and show that this novel C-terminal domain does not exhibit any detectable hydrolytic or oxidase activity but independently interacts with soluble and resistant starches. This novel SBD constitutes a new CBM family that assists large and multi-modular α -amylases in forming pores in resistant starch granules. Based on amino acid sequence similarity searches, we identified that this novel domain is most often associated with large (>1000 aa) and multi-modular GH13 α -amylases that contain additional starch binding CBMs (CBM25 and CBM26) and several FNIII domains. Interestingly, Domain 2 most often occurs in Bifidobacteria species that are associated with the human gastrointestinal tract; therefore it is most likely important in facilitating resistant starch degradation by bacteria in the mammalian gut.

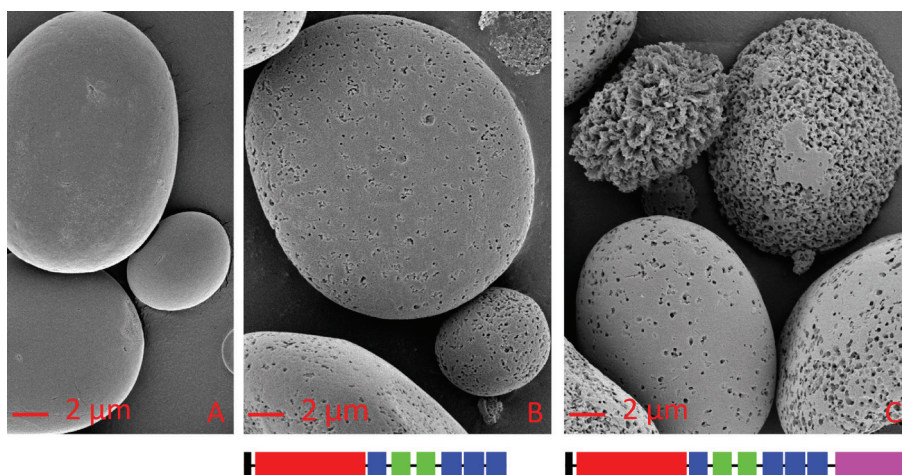


Figure 1: SEM images of wheat starch granules. **A:** incubated for 48 h with negative control sample (empty vector sample), **B:** incubated for 72 h with MaAmyA7, **C:** Incubated for 72 h with MaAmyA. The domain organization of the enzyme used for incubation of the granules is show below the images, using the following colors to indicate the different domains; ■: signal sequence; ■:GH13 catalytic domain; ■: FNIII domain; ■:CBM25 domain; ■:Domain 2. For more details see chapter 2, (published as [160]).

Materials and methods

Bioinformatic tools

All BLAST searches were performed with NCBI BLASTP using standard settings. Conserved domains were detected using both the NCBI conserved domain finder [135] with forced live search, without low-complexity filter, using the conserved domain database (CDD) and dbCAN [26] with standard settings. Alignments were made with Mega6.0 [137] using its build-in muscle alignment with standard settings and manually tuned. Alignments were visualized with Jalview 2.8.1 [166]. Phylogenetic trees were made with Mega6.0 using maximum likelihood method with gaps/missing data treatment set on partial deletion instead of full deletion. Trees were visualized with Interactive Tree Of Life v2 [138]. Information about the GH13 subfamilies was obtained from the CAZy database [20]. The domain organization shown in the tree is based on the combined dbCAN and CDD data. Signal sequences were predicted with SignalP 4.1 using “Gram-positive bacteria” organism group and “Sensitive” D-cutoff values [136]. Domain prediction servers SBASE, DOBO and DOMpro [167-169] were used with standard settings. Secondary and tertiary structure prediction was done using the Phyre2 server with standard settings [29].

Cloning and expression

CBM74 (aa1116-1415) was cloned from the *M. aurum* B8.A *amyA* gene construct (chapter 2, published as [160]) into pET15b using the LIC system [170,171] and the Fwd and Rev primers CAGGGACCCGGTGCGCTCTACTCGACCAACCCGTCGTCGAG and CGAGGAGAAGCCCGGTTACAAGAAGCCTACGCTCGCGAAGCGAGC. A recombinant *E. coli* strain with an empty pET15b vector was used to produce a negative control sample.

Production and purification of CBM74 protein

CBM74 -pET15b was transformed into *E. coli* BL21* DE3 cells (Novagen). One liter of LB broth supplemented with ampicillin (50 ug/ml) was inoculated with a 5 ml overnight starter culture and incubated with shaking at 37 °C until an OD₆₀₀ of 1.0 was reached. Protein production was induced by the addition of 1 mM IPTG and the culture was incubated for a further 16 h at 20 °C. *E. coli* cells were harvested by centrifugation (4,250 g for 20 min) and the pellet was resuspended in 25 ml of 20 mM Tris-HCl pH 8.0, 500 mM NaCl (Buffer A) containing 0.2 mg/ml lysozyme and 0.2 mg/ml DNaseI and lysed by sonication. The lysed cells were subjected to centrifugation (15,000 g for 45 min). The supernatant did not contain any soluble protein, CBM74 protein was only present in inclusion bodies in the cell pellet. The cell pellet was washed twice with Buffer A containing 0.1% Triton X-100 to remove membrane debris, then washed twice with Buffer A. The inclusion bodies were then denatured by resuspension in 200 ml Buffer A containing 8 M urea, and stirred overnight at room temperature. The following day the solution containing denatured CBM74 protein was spun down (4,250 g

for 20 min) and the denatured protein was dialyzed in a stepwise manner into 2 L of Buffer A supplemented with 5 mM CaCl_2 and 5 mM MgCl_2 and decreasing concentrations of urea (4 M, 2 M, 0 M). Each dialysis step was performed over 24 h, and the final 0 M urea dialysis step was performed twice over 48 h using snakeskin dialysis membrane with a 10 kDa MWCO pore size (Thermo Scientific). The resulting dialyzed supernatant contained soluble, refolded CBM74 protein which was >95% pure as assessed by SDS-PAGE (not shown), and yielded >300 mg soluble CBM74 protein per liter of *E. coli* culture. CBM74 protein was further purified by immobilized metal affinity chromatography (IMAC), taking advantage of the N-terminal His_6 tag present. The IMAC purification was carried out using established protocols (chapter 2, published as [160]). Purified proteins were stored at 4 °C in 50 mM Tris-HCl buffer pH 6.8 containing 10 mM CaCl_2 .

Granular starch binding

All binding studies were performed in standard binding buffer (50 mM Tris-HCl buffer pH 6.8 containing 10 mM CaCl_2). Granular wheat (Sigma-Aldrich, catalog no. S5127), waxy corn (Sigma-Aldrich, catalog number S9679) and potato starch (AVEBE), and cellulose (Sigma-Aldrich catalog number C6413) were washed with standard binding buffer and 0; 0.05; 0.1; 0.25; 0.5; 0.75; 1.0; 1.5; 2.0; 2.5; 5.0; 7.5% (m/v) suspensions of granules were prepared in the same buffer. Of each suspension, 100 μl was transferred to a clean 2 ml reaction tube and the buffer removed through centrifugation (5,000 g for 20 sec). Subsequently, 100 μl His-Tag purified CBM74 (0.3 mg/ml), 100 μl of empty vector negative control sample or 100 μl of BSA (0.3 mg/ml) was added to each pellet (in triplicate). For cellulose and BSA, only suspensions of 2.5% (m/v) were included. The mixtures were incubated for 2 h at 4 °C on a roller bench. Unbound protein was removed by centrifugation (10,000 g for 15 sec). Pellets of the 5% (m/v) samples were kept for additional experiments. The supernatant was transferred to a microtiter plate suitable for UV measurements (Falcon), and assayed at 280 nm in a microtiter plate reader (Spectramax Plus; Molecular Devices, Sunnyvale, CA) with a path length of 0.25 cm.

Pellets of the 5% (m/v) wheat and potato starch granule suspensions were washed three times with 100 μl standard binding buffer for 30 min at 4°C, followed by centrifugation; the supernatant of the 3rd wash was collected. Each 100 μl suspension was then split into two 50 μl suspensions, resulting in six 50 μl suspensions for each granule type. Two elution steps for bound CBM74 were performed. In the first step 50 μl standard binding buffer containing 5% (m/v) of the carbohydrate (buffer, maltose, glucose, dextrose, iso-maltose or mannose) to be tested for elution was added, mixed for 30 min on a roller bench at room temperature, and collected by centrifugation. In the 2nd elution step, 50 μl 5x concentrated SDS sample buffer was added, mixed for 5 min and granules collected through centrifugation. Of the third washing and first elution steps, supernatant fractions of 20 μl were mixed with 5 μl 5x SDS sample buffer and

loaded onto SDS-PAGE. Of the 2nd elution 20 µl supernatant was mixed with 5 µl buffer and loaded onto SDS-PAGE; on each gel a protein marker (Fermentas) and a negative control (empty vector) sample were also included. Afterwards the gels were stained with Coomassie Brilliant Blue R (Bio-Rad) to visualize the protein bands, or used for semi-dry Western blot (Bio-Rad). Additional controls were performed to exclude any effects of proteins naturally attached to starch granules (washed granules eluted with SDS-sample buffer).

The amount of CBM74 bound to the starch granules was determined with the calculated molar extinction coefficient (using the ExPASy ProtParam tool [172]) of CBM74 (49850 M⁻¹·cm⁻¹) and the following formula:

$$\text{Bound CBM} = \frac{A_{280 \text{ total}} - A_{280 \text{ unbound}}}{\epsilon \cdot l}$$

Where:

Bound CBM	= Concentration of CBM74 that bound to the granules (M)
$A_{280 \text{ total}}$	= Absorbance of total CBM74 protein available for binding at zero time
$A_{280 \text{ unbound}}$	= Absorbance of unbound CBM74 protein after 2 h incubation with granules
ϵ	= Calculated molar extinction coefficient of CBM74 (= 49850 M ⁻¹ ·cm ⁻¹)
l	= Spectrophotometer path length (= 0.25 cm)

Using a Scatchard plot [173,174] the concentration of bound CBM74 divided by the concentration of unbound CBM74 was plotted against the concentration of bound CBM74. All concentrations were normalized to an equal amount of starch granules.

The dissociation constants (K_d) were determined though non-linear regression analysis with Microsoft Excel 2010 as described by Kemmer et al. [175], using a one site binding model [176]:

$$\text{Bound CBM} = \frac{B_{\text{Max}} \cdot [S]}{K_d + [S]}$$

Where:

Bound CBM	= Concentration of CBM74 that bound to the granules (M)
B_{max}	= The maximum binding capacity
$[S]$	= Starch granule concentration (mg/ml)
K_d	= Dissociation constant (mg/ml)

The K_a values are equal to K_d^{-1}

Western blot

Samples were transferred onto a nitrocellulose membrane (Pharmacia) through semi-dry blotting (Trans-Blot Semi-Dry SD cell, Biorad) for 15 min at 20 V, using transfer buffer (50 mM Tris-HCl, 40 mM Glycine, 1.75 mM SDS, pH 9). After blotting, membranes were blocked for 1 h at room temperature with blocking buffer (140 mM NaCl, 10 mM phosphate buffer, and 3 mM KCl, pH 7.4) (Calbiochem, PBS tablets) containing 1% m/v BSA (Sigma) and 0.05% Tween 20 (Sigma). Then the membranes were incubated for 1 h with blocking buffer containing 0.02% v/v 1-step Anti-His antibody (Qiagen), and washed (with block buffer) 3 times for 5 min. After washing, membranes were activated with fresh mixed ECL reagent (Pharmacia) and exposed in a Chemidoc (Bio-Rad) for up to 30 min.

Polysaccharide macroarray binding analysis

The macroarray method used is based on the procedure described in [177]. All carbohydrates were obtained from Sigma unless indicated otherwise. Soluble potato starch, granular potato starch (AVEBE), granular wheat starch, granular waxy corn starch, amylopectin, maltodextrin and pullulan were dissolved in Milli-Q at a concentration of 10 mg/ml (m/v). Granular starches were (partially) dissolved by heating the suspension in a heating block set at 100 °C for 10 min. Amylose was dissolved by adding 0.1 M NaOH and subsequent addition of an equal volume of 0.1 M HCl. Macroarrays were prepared by spotting 1 µl of each dissolved carbohydrate onto a nitrocellulose (Pharmacia) membrane. The His-Tagged protein MaAmyA7 lacking domain CBM74 and containing both an N- and C-terminal His-Tag (chapter 2, published as [160]) was used as a positive method control for proper His-Tag detection on each membrane. After spotting, membranes were dried to the air for at least 2 h. After blocking (see Western blot) the membranes were probed with 300 µg His-Tag purified CBM74, 100 µg of His-Tag purified CBM41 (a well-characterized starch binding domain) as a positive control [177] or an equal volume of the negative control (empty vector) sample in 10 ml blocking buffer and incubated at 4 °C for 1 h. Subsequently membranes were treated as regular Western blots after blocking.

CBM74 enzyme catalytic activity testing

The potential catalytic activity of CBM74 on starch was tested by incubation of 30 µg CBM74 in 1 ml standard binding buffer (50 mM Tris-HCl buffer pH 6.8 containing 10 mM CaCl₂) containing 10 mg/ml soluble potato starch. To test for LPMO activity [6] the following was added: 10 mM CuCl₂; 5 mM L-cysteine (adjusted to pH 6.8); 5 µl (30 unit) barley β-amylase (Megazyme, Ireland, 1,000 fold diluted in standard binding buffer) as well as all possible combinations of these additions. Reactions were incubated for 24 h at 37 °C in a heating block. Products formed were analyzed on TLC as described previously [178] and MALDI-TOF MS (Shimadzu AXIMA Performance) using 2,5-dihydroxybenzoic acid (DHB) as matrix. All incubations and analyses were performed in duplicate.

Results

Identification of a novel protein domain in MaAmyA

Recently we reported the characterization of MaAmyA, a large and multi-domain α -amylase from *M. aurum* B8.A (1417 aa) which is able to form pores in starch granules. MaAmyA carries 2 CBM25 domains and 4 FNIII domains, plus a novel protein domain at its C-terminus that is approximately 300 aa (Domain 2) (chapter 2, published as [160]). Deletion of this Domain 2 did not affect the ability of MaAmyA7 to hydrolyze granular starch, but the average pore sizes in starch granules were reduced 3-fold (Fig.1). This strongly suggests that Domain 2 has a specific functional role, which is investigated here both experimentally and with various bioinformatics tools.

The 300 aa C-terminal tail encoding the predicted Domain 2 of MaAmyA (Genbank AKG25402.1, aa1116-1415) was successfully cloned and expressed in *E. coli*. Most Domain 2 protein accumulated in inclusion bodies. After denaturing and refolding, soluble Domain 2 protein was obtained. SDS-PAGE analysis revealed a protein of 37.5 kDa, matching the predicted size of Domain 2, with > 85% purity (based on SDS-PAGE analysis) (data not shown).

Domain 2 of MaAmyA binds to soluble and insoluble starches

The Domain 2 protein (7 μ M) was able to bind to wheat, potato and waxy corn starch granules present in a 5% m/v suspension. The effects of various carbohydrates on this binding of Domain 2 (7 μ M) to the granules was studied. After Domain 2 binding, starch granules were washed and elution was attempted with 5% (m/v) solutions of maltose, glucose, dextrose, iso-maltose or mannan. None of these carbohydrates elicited the release of a detectable amount of Domain 2 from the granules (using SDS-PAGE or Western blot analysis with anti-His-Tag). In a second elution step SDS-sample buffer efficiently released bound Domain 2 from the starch granules. SDS-PAGE analysis of the samples obtained showed protein bands corresponding to the expected mass of Domain 2 protein at 37.5 kDa. Also Western blot analysis with anti-His-Tag antibodies showed a single band at 37.5 kDa. An empty vector negative control sample did not show any bands (data not shown).

To examine the possible interactions of Domain 2 with non-granular carbohydrates, a macroarray containing starches from several sources, as well as various other polysaccharides, was prepared. Purified His-Tagged Domain 2 protein was allowed to bind to the nitrocellulose-bound starches, and its binding was visualized using anti-His-Tag antibodies (Fig. 2). Domain 2 was shown to bind to all tested starches, plus amylose and amylopectin. No relevant signals were detected in the empty vector negative control sample. No binding of Domain 2 to any of the non-starch polysaccharides was observed. Positive controls included on each macroarray yielded expected results. Domain 2 of MaAmyA

thus represents a novel CBM, an SBD that is able to bind to amylose, amylopectin and starch granules.

To determine the affinity of Domain 2 for starch binding, 7 μ M Domain 2 protein was incubated with increasing percentages of different types of starch granules (Fig. 3). Microgranular cellulose (2.5% m/v) was included as a negative control and did not show any interaction with Domain 2 (data not shown). At low concentrations of starch granules a clear relation was observed between the amount of Domain 2 bound and the concentration of starch granules. At higher granule concentrations the amount of bound Domain 2 leveled off, indicating that all Domain 2 that was able to bind (45-75% of total) after the denaturation/renaturation isolation procedure from inclusion bodies had bound to the starch granules. Potato starch showed a different pattern compared to wheat and corn starch. Potato starch saturation was reached at a lower concentration of starch granules and the total amount of Domain 2 that bound was significantly lower than with wheat and waxy corn starch (Fig. 3). The Scatchard plots [173] in which the concentration of bound Domain 2 divided by unbound Domain 2 was plotted against the concentration of bound Domain 2 were linear for all three starch types which suggests a single mode of binding for Domain 2 (no cooperativity). When it is assumed that Domain 2 has a single mode of binding, the estimated K_a values are: 0.15 ± 0.02 mg/ml for wheat starch granules; 0.14 ± 0.02 mg/ml for waxy corn starch granules and 1.4 ± 0.5 mg/ml for potato starch granules. These affinities are in the same range as reported for other SBD such as CBM20 and CBM41 [132,177]. An empty vector negative control series did not show any protein binding with any of the granules. In addition, no aspecific protein binding to the granules was observed with bovine serum albumin (BSA).

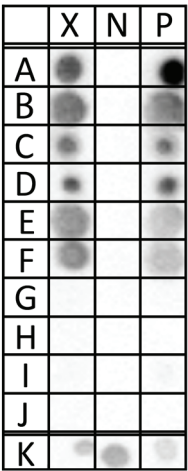


Figure 2: Polysaccharide binding macroarray with detection of bound His-Tagged CBM74 and CBM41 proteins.

Polysaccharide substrates (rows):
A: soluble potato starch
B: boiled granular potato starch
C: boiled granular wheat starch
D: boiled granular waxy corn starch
E: amylose
F: amylopectin
G: pullulan
H: dextran
I: glycogen
J: cyclodextrin
K: N- and C-terminally His-Tagged MaAmyA7 protein (method positive control)

Protein samples (columns):
X : CBM74
N: negative control (empty vector sample)
P: positive control (His-Tagged CBM41 sample) [177]

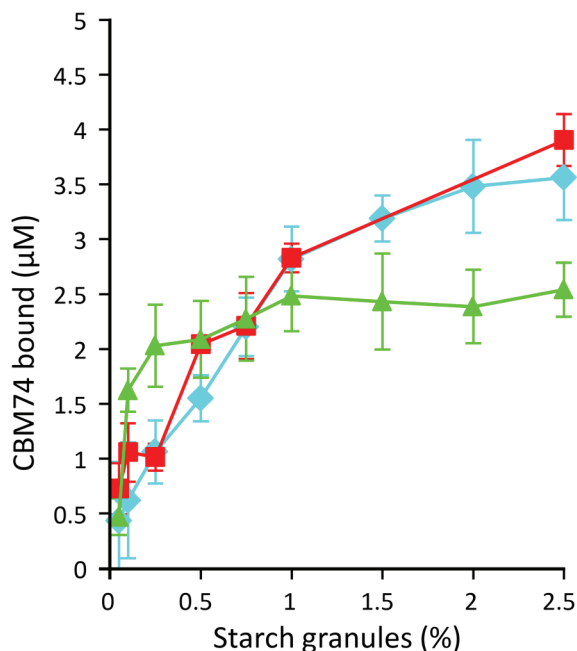


Figure 3: Binding of CBM74 protein (7 μ M) to increasing amounts of different starch granules (0-2.5%, m/v). The unbound CBM74 concentration after incubation with starch granules was determined in triplicate through UV measurements at 280 nm, using the calculated molar extinction coefficient of CBM74. The concentration of bound CBM74 was then calculated by subtracting the value after binding from an included total protein value (without starch granules). CBM74 did not bind to 2.5% m/v microgranular cellulose while BSA did not show binding to any of the granules tested (data not shown). Values are means \pm SD.

◆ : wheat starch granules
 ■ : waxy corn starch granules
 ▲ : potato starch granules

Domain 2 does not show hydrolytic or LPMO enzymatic activity

Domain 2 was tested for enzyme catalytic activity on soluble potato starch. Since a eukaryotic starch-degrading lytic polysaccharide monooxygenase (LPMO) was recently discovered that needed a cofactor (cysteine) and β -amylase to visualize its activity [158,159], we performed similar co- incubations with Domain 2 protein. Even after incubation of 30 μ g of Domain 2 for 24 h, no products were detected using TLC and MALDI-TOF MS analysis. Under the conditions tested we thus were unable to detect any starch acting hydrolytic or LPMO activity for Domain 2.

Occurrence of Domain 2 in bacterial genomes

A BLAST search with the Domain 2 amino acid sequence returned 77 hits (November 2015) for a 286-328 aa long fragment ($E > 4 \cdot 10^{-25}$), all from bacterial origin. Three additional significant hits were ignored as these sequences were incomplete; in all cases the partial domain was located adjacent to a gap in the genome sequence. Domain 2 thus is not unique for MaAmyA and occurs more widespread in bacterial proteins. In view of its starch binding activity, absence of enzymatic activity, stimulatory effect on pore formation by the MaAmyA enzyme, and more abundant distribution in bacteria, we conclude that Domain 2 proteins constitute a novel CBM family, designated family CBM74.

For secondary and tertiary structure prediction the amino acid sequence of the CBM74 domain of MaAmyA was submitted to the Phyre2 server [29]. The predicted structure revealed no similarity to any known structures. Only a

fragment (aa 14-110) showed resemblance to the structure of CBM9 in Xylanase A of *Thermotoga maritima* MSB8 (PDB:1I8A[A] (78% confidence) although the amino acid identity is low (24%) [179]. CBM9 has only been found in association with xylanases [180]. The predicted structure for this part of CBM74 showed 5 β -sheets with high confidence scores, similar to CBM9. The remaining aa 111-300 fragment did not show significant structural similarity with other known proteins. The predicted structure for this part of CBM74 showed 7 additional β -sheets with high confidence scores (Fig. 4).

Alignment of CBM74 of MaAmyA and its 76 homologs

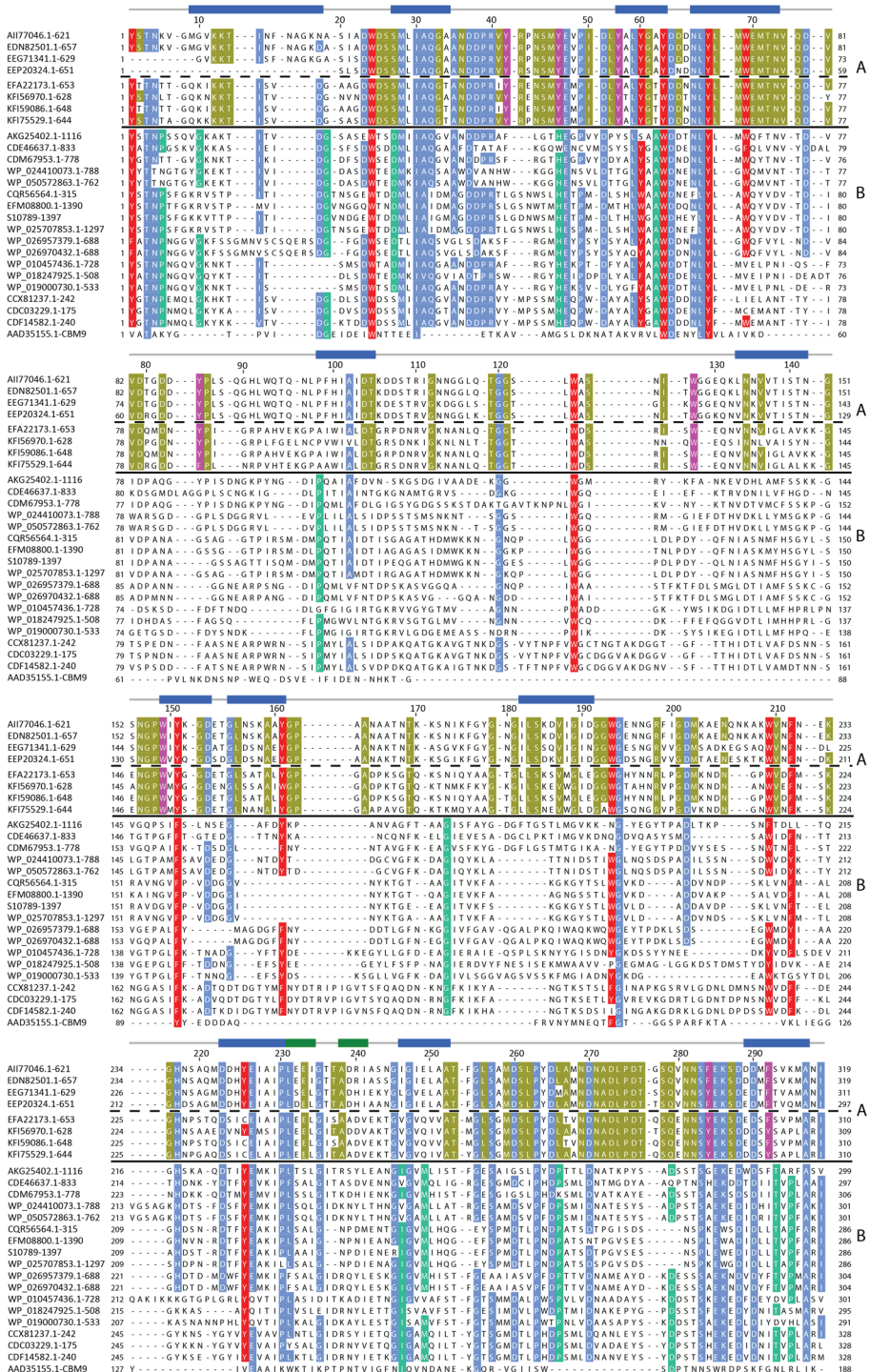
A sequence alignment was made with all CBM74 homologs (Fig. 4). MaAmyA CBM74 has 34-60% identity and 48-73% similarity with its 76 homologs. Several conserved aromatic residues were identified, which may be of special interest since these are often involved in carbohydrate interaction and binding [46,181,182]. The overall similarity is lower in the middle part of the CBM74 domain (aa113-201) although some aromatic residues are conserved here as well. Based on the alignment, two clusters were defined: cluster A with 50 sequences (45-67% identity, 61-78% similarity) and cluster B (35-60% identity, 48-73% similarity) with 27 sequences, including CBM74 from MaAmyA (see also the phylogenetic tree in Fig. 5). The homologs in cluster A contain an additional 86 conserved residues, including 8 aromatic residues, compared to the homologs in cluster B.

CBM74 is a single domain protein

The results of the Phyre2 prediction showed structural similarity between aa14-110 of CBM74 and CBM9. When CBM9 of Xylanase A from *T. maritima* MSB8 (GenBank AAD35155.1) was included in the alignment, the first tryptophan (aa 72) of CBM9, which is known to be involved in ligand binding [179], aligned with the conserved tryptophan at aa 70 in cluster B (Fig. 4). Within cluster A, this tryptophan is mostly substituted by a tyrosine. The second tryptophan (aa176) involved in ligand binding in CBM9 is not conserved in CBM74. CBM74 is much larger than CBM9 and has multiple other conserved aromatic residues that may be involved in further interactions with starch.

Figure 4 (next page): Sequence alignment of all 77 known CBM74 homologs (November 2015), including MaAmyA CBM74. CBM9 (aa 729-843) of Xylanase A of *Thermotoga maritima* MSB8 was included as reference [179]. For visibility, only a limited number of the CBM74 sequences (8 out of 50 of cluster A and 17 out of 27 of cluster B) are shown representing the maximal amount of variation visible between the sequences. The aa numbers are based on MaAmyA CBM74. The solid black line indicates the separation between clusters A and B. The dashed black line indicates the 2 subgroups in cluster A. The gray line shows the secondary structure of CBM74 from MaAmyA (aa 1116-1416 of AKG25402.1) as predicted by Phyre2, with blue boxes representing β -sheets and green boxes representing α -helices. Color code used for amino acids: ■: aromatic residues conserved in all sequences; ■: aromatic residues only conserved in cluster A; ■: other conserved residues in all sequences; ■: only conserved in cluster A; ■: only conserved in cluster B. GenBank accession numbers are used as names followed by the aa number where the similarity with MaAmyA CBM74 protein starts.

CBM74 is a novel starch binding domain



The similarity between CBM9 and only the first part of CBM74 may indicate that CBM74 is in fact a combination of 2 domains. To investigate whether CBM74 represents one or two domains, the aa sequence of the CBM74 domain of MaAmyA was submitted to three domain prediction servers (SBASE, DOBO and DOMpro) [167-169,183], which all predicted that it represents a single domain. Also in view of the observation that all 77 CBM74 homologous domains in the databases have a similar length (286-328 aa), we conclude that CBM74 represents a single domain protein.

Phylogenetic analysis of all family CBM74 members

A phylogenetic tree based on the alignment of all family CBM74 homologs and a selection of known CBM sequences from CAZy is shown in Figure 5, along with the domain organization of the proteins they belong to. The phylogenetic tree shows that all CBM74 homologs cluster together as a new group, separate from previously described CBMs. The CBM74 homologs are most closely related to CBM9, also reflecting the structural homology described above. Family CBM74 shows clustering that in general matches with the host species that harbors the CBM74 containing protein (Fig. 5). Clusters A and B, as identified in the sequence alignment (Fig. 4) are clearly visible in the tree as well. Cluster A is the largest CBM74 cluster, containing all the CBM74 members that are part of proteins from mainly *Bifidobacterium* species, while cluster B consists of all the others (Fig. 4, 5). CQR56564 is most likely linked to the α -amylase that is preceding it in the genome (CQR56565.1). Therefore CQR56564.1 was linked to this α -amylase to reveal the full organization of the protein. For comparison both are shown in the phylogenetic tree (Fig. 5).

Of the 77 unique (and restored) CBM74 proteins, 69 CBM74 domains are part of Glycoside Hydrolase family 13 (GH13) α -amylases. They all have the ABC-domains typical for α -amylases [184], even though this is not always shown in the domain organization (Fig. 5). This is due to the fact that a number of these proteins possess C-domains with a primary sequence that has a low identity with C-domains currently in databases. However, structural analysis using Phyre2 [29] revealed that despite this low sequence identity all proteins shown have a predicted fold that is similar to that of GH13 C-domains, including the all β -sheet fold and typical greek key motive [184]. Of all 77 CBM74 containing proteins, only MaAmyA of *M. aurum* B8.A has been characterized experimentally, namely as a granular starch degrading α -amylase (chapter 2, published as [160]). Most of the 69 CBM74 containing α -amylases have catalytic domains that belong to the GH13_28 subfamily (59 sequences) or GH13_19 (9 sequences). Only MaAmyA from *M. aurum* belongs to the GH13_32 family (chapter 2, published as [160]). The CBM74 domain is generally present in the middle (64 sequences) of the protein, but never directly adjacent to the catalytic domain (Fig. 5). It is also found at the C-terminus of these proteins (13 sequences), but never at the N-terminus. MaAmyA has a C-terminal CBM74 which is preceded by 3 FNIII

domains (chapter 2, published as [160]). FNIII domains are only found in 8 other CBM74 containing proteins; these proteins mostly contain a GH13_19 catalytic domain, one or more CBM25 domains and a C-terminal CBM74 (Fig. 5). Of the CBM74 proteins, 8 are not linked to a catalytic domain. In most cases, however, α -amylase catalytic domains are encoded by immediately adjacent genes in their respective genomes.

CBM25 and the structurally related CBM26 domain [133] are commonly present in the CBM74 containing α -amylases (Fig. 5). At least one CBM25 (in 25 sequences) or CBM26 (in 43 sequences) domain is present about 150 aa after or about 200 aa (300 aa only in case of MaAmyA) before CBM74 in these α -amylases. In 5 out of the 8 CBM74 containing proteins without a catalytic domain, CBM26 is present about 200 aa before the CBM74 domain. The general domain organization of CBM74 containing proteins, and the location of this domain in these proteins, appears to be related to the identity of the bacterial host species (Fig. 5) (see next paragraph).

Bacterial species harboring proteins with CBM74 homologs

Our data show that 69 of the 77 CBM74 homologs are present in large and multi-domain putative α -amylases that are mainly encoded by bacteria isolated from the mammalian gut or gut related environments. The 50 proteins with CBM74 homologs in cluster A mainly originate from *Bifidobacterium* species (48 proteins) while 2 originate from *Prevotella* species. Most of these species were isolated from mammalian gastrointestinal tract (GIT) related environments (45 species), 2 were isolated from hamster dental plaque and 1 from chicken GIT, while for 2 the source of isolation is unknown (Fig. 5). All CBM74 containing proteins in this cluster are large and multi-domain α -amylases that belong to the GH13_28 subfamily.

The CBM74 containing proteins from *Bifidobacterium* can be split into 2 groups; with and without CBM26. The general domain organization for the group with CBM26 is: GH13_28 catalytic domain, ~150 aa gap, CBM74, Big_2, CBM26, Big_2, additional binding domains (CBM13, 20 or 25). Big_2 is a bacterial domain with an Ig-like fold, commonly found in bacterial and phage surface proteins [135]. The Big_2 domain is widely distributed in carbohydrate acting enzymes. Its function is not clear, but removal of the Big_2 domain from a termite gut bacterium GH10 xylanase greatly reduced the activity of this enzyme [185]. In a few cases CBM26 is replaced by CBM25. The general domain organization for the group without CBM26 is: GH13_28 catalytic domain, ~150 aa gap, CBM74, Big_2, CBM25, 1-3 Big_2, 1-3 SLH (Surface Layer Homology) domains. Some shorter members lack the SLH domains. The C-terminal SLH domains are associated with non-covalent anchoring to the cell surface S-layer via a conserved mechanism involving wall polysaccharide pyruvylation [186,187]. Interestingly this group without CBM26 forms a separate subgroup within CBM74 cluster A (Fig. 5). The two proteins from *Prevotella* species are shorter and consist of a GH13_28 catalytic domain, a CBM26 domain, and a C-terminal CBM74 domain.

CBM74 is a novel starch binding domain

Legend protein domain organization:

◆ CBM74

Shapes:

□ Complete domain

◐ Incomplete domain

◑ Domain incomplete N-terminus

◒ Domain incomplete C-terminus

◓ Position in tree based on aa Sequence of this domain

Domain colors:

■ GH10 catalytic region

■ GH14 catalytic region

■ GH13 amylase AB region

■ GH13 amylase C region

■ FNIII

■ SLH

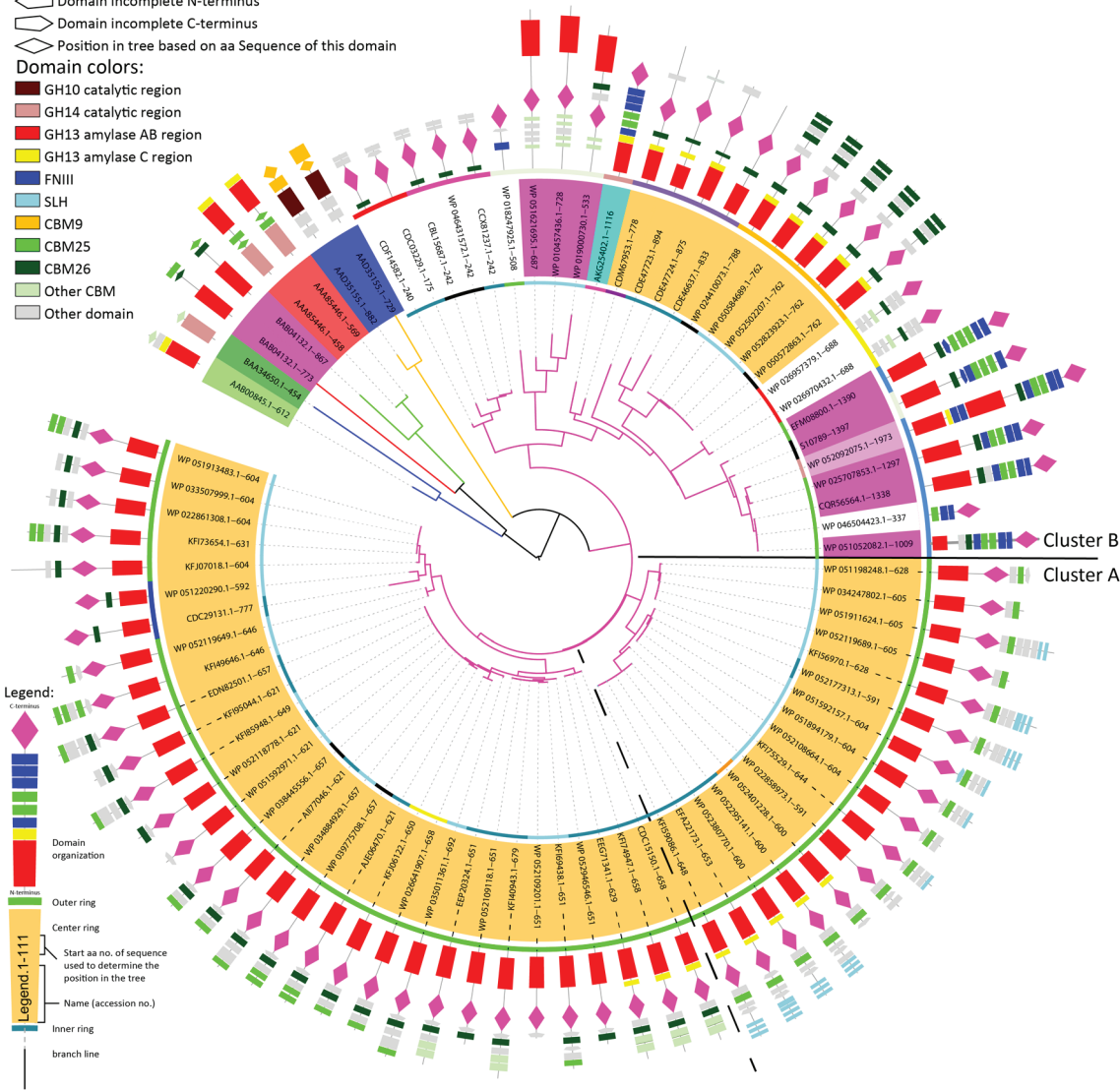
■ CBM9

■ CBM25

■ CBM26

■ Other CBM

■ Other domain



Branch line colors:
indicating CBM
family clustering

■ CBM9
■ CBM20
■ CBM25
■ CBM26
■ CBM74

Inner ring colors: indicating the
isolation source of the bacteria
containing the protein

■ Mammalian gut related
■ Bird gut related
■ Human gut related
■ Dental plaque
■ Pasteurized milk
■ Soil
■ Seawater
■ Potato plant waste water
■ Scale biogas reactor
■ No Information

Center ring colors:
GH (Sub) family of
the catalytic domain

■ GH10
■ GH13_2
■ GH13_5 & GH13_19
■ GH13_19
■ GH13_28
■ GH13_32
■ GH13 & GH14
■ GH14
■ No catalytic domain

Outer ring colors : indicating
the species of the bacteria
containing the protein

■ *Alliagarivorans*
■ *Bifidobacterium*
■ *Clostridium*
■ *Eubacterium*
■ *Microbacterium aurum*
■ *Paenibacillus*
■ *Prevotella*
■ *Ruminococcus*
■ *Streptococcus*
■ Other

— 0.1

Figure 5 (previous page): Phylogenetic tree of all 77 known CBM74 homologs (November 2015), including MaAmyA CBM74, together with a selection of sequences from CBM9, CBM20, CBM25 and CBM26 for which the 3D structures are known (based on <http://www.CAZy.org>). The part of the full protein sequence that was used to construct the tree (the CBM74 domain) is shown as a diamond. The domain organization of the full proteins shown is based on combined CDD, DBcan and Blast (for CBM74) data. The CBM74 domain is indicated in pink. CDD does not recognize C-domains in all α -amylases. However, manual assessment of the amino acid sequence downstream of the indicated GH13 AB-domain using Phyre 2 revealed that a C-domain is present in all proteins (see text). The outer ring between the accession numbers and the protein domain organization indicates the bacterial host species, while the inner ring between the tree and the accession numbers shows the source of isolation of the host organism. Tree line colors correspond to the different CBM families. The background color of the accession numbers indicates the GH (sub)family of the catalytic α -amylase domain. The solid line shows the separation between CBM74 cluster A and B. The dashed line shows the separation between the 2 subgroups of cluster A. The scale bar indicates 0.1 amino acid replacement per site.

The 27 proteins with CBM74 homologs in cluster B have more diverse origins (Fig. 5): 5 *Paenibacillus* strains, all isolated from soil (5 strains); 5 *Streptococcus* strains, mainly isolated from mammalian gut related environments (3 strains); 4 *Clostridium* strains, mainly isolated from mammalian gut related environments (3 strains); 2 *Eubacterium* strains from mammalian gut related environments; 2 *Aliagarivorans* strains isolated from seawater; 1 *M. aurum* strain (studied in this paper) from a potato waste water treatment plant [127]; 3 *Ruminococcus* strains of which 1 from mammalian gut related environments; 1 *Ruminobacter* strain from a mammalian gut related environment; 1 *Succinivibrionaceae* strain from mammalian gut related environments; 1 *Succinimonas* strain from mammalian gut related environment and 1 *Orenia marismortui* strain isolated from soil. The five CBM74 containing proteins from *Paenibacillus* strains stand out as these are the only ones next to MaAmyA that contain FNIII domains, CBM25 domains and a C-terminal CBM74 (Fig. 5). Despite this similar domain organization, the individual CBM25 and FNIII domains of MaAmyA do not show high similarity with those from the *Paenibacillus* enzymes or with the CBM25 domains from other CBM74 containing enzymes in phylogenetic trees based on either domain (data not shown).

Within the group of CBM74 containing proteins, specific GH13 α -amylase subfamilies can be linked to different bacterial species (Fig. 5). For example, CBM74 containing α -amylases that belong to GH13_28 are found in *Clostridium* (4), *Bifidobacterium* (48), *Streptococcus* (5) strains and *Prevotella* (2), while those belonging to GH13_19 are found in *Paenibacillus* (5), *Eubacterium* (1), *Ruminobacter* (1), *Succinivibrionaceae* (1) and *Succinimonas* (1) strains. MaAmyA of *M. aurum* is the only GH13_32 protein (chapter 2, published as [160]).

Discussion

The large and multi-domain MaAmyA α -amylase from *M. aurum* B8.A (1417 aa) is able to degrade granular starch (Fig. 1) and contains a novel domain at its C-terminus. This 300 aa Domain 2 is able to bind to raw starch granules (Fig. 3) as well as to amylose and amylopectin (Fig. 2). The length of Domain 2 is comparable to the length of the recently described starch degrading LPMO [158,159]. Since one LPMO family, now defined as Auxiliary Activity 10 (AA10), was initially characterized as a CBM (CBM33) [188] we screened Domain 2 for mono-oxygenase activity but were unable to find any. Interestingly, currently (Feb 2016) identified LPMOs, defined as AA families 9, 10, 11 and 13 in the CAZy.org database, do not contain any additional catalytic domains and are part of relatively small proteins (average 350 aa) that usually contain no more than two additional domains [20]. This is unlike Domain 2 which is usually part of large multi-domain proteins containing a GH13 catalytic domain. This sets Domain 2 apart from currently known LPMO's. Since Domain 2 is usually found combined with a GH13 catalytic domain, a non-catalytic function seems more likely for Domain 2. A majority of the Domain 2 containing proteins have a predicted signal sequence and are therefore likely secreted by the host, Domain 2 could also act as a cell wall anchoring domain. However, such domains are usually located at the protein termini [186,187,189,190], while Domain 2 is often found in the middle. It therefore seems unlikely that Domain 2 functions as a cell wall anchoring domain.

In previous work (chapter 2, published as [160]) the full length MaAmyA enzyme and a mutant with deleted Domain 2 (MaAmyA7) showed similar starch degrading activity with both soluble and granular starch. As a major difference, the pores formed in starch granules by MaAmyA were about three times larger than those formed by MaAmyA7 (chapter 2, published as [160]). These results suggest that Domain 2 plays a specific role in binding to starch granules (Fig. 2, 3), thereby assisting in their degradation (Fig. 1).

No specific enzyme activity was found associated with Domain 2 itself. This MaAmyA Domain 2 thus appears to constitute a novel SBD/CBM, and was designated CBM74. It displays highest affinity for binding to potato starch granules (Fig. 3). Although potato starch granules are larger than wheat and maize starch granules [176], this does not automatically result in a higher affinity. In a binding study of Pig Pancreatic Amylase (PPA) binding to different starches granule types it was shown that PPA had a lower affinity for potato starch than maize and wheat starch. In addition when one type of starch granules was separated into two pools based on the granule sizes, PPA showed higher affinity for the smaller granules [176]. In a study with CBM20 it was found that the affinities for potato and maize starch granules were similar [191]. The differences in affinity could also be related to the crystallinity type of the starch granules, which is mainly

dependent on the plant species that produced the granules [192]. Since potato starch granules have a B-type crystallinity while wheat and maize starch granules have an A-type crystallinity [192], this corresponds with the differences in affinities we found. This could indicate that CBM74 has a higher affinity for B-type crystallinity granules. However more research is needed to fully understand the mechanism of binding of CBM74.

CBM74 is 300 aa long and therefore exceptionally large compared to other known CBMs which are generally between 50 and 200 aa long [20,193]. It is noteworthy that 90% of all identified protein domains are shorter than 200 aa [194,195]. Several domain prediction servers indicated CBM74 to be a single domain. All 77 CBM74 homologs identified in the present study have a similar length and showed similarity over the full ~300 aa, thus also indicating that CBM74 is a single and complete domain without internal duplications. Therefore we conclude that CBM74 is indeed a single domain and an extraordinarily large CBM.

CBM74 clearly occurs more widespread and is commonly part of extremely large (>1300 aa) multi-domain GH13 amylases that also contain CBM25 or CBM26 domains next to a single catalytic domain. Less than 2% of all GH13 members currently listed in the CAZy database are 1300 aa or longer [20]. On average GH13 α -amylases are about 650 aa long with usually only up to two additional domains [20]. The α -amylase proteins with a CBM74 domain appear to be specialized in the degradation of starches that are difficult to hydrolyze enzymatically.

Most of the currently known CBM74 containing α -amylases (at least 80%) originate from bacteria isolated from the GIT (Fig. 5). This number could be slightly biased due to the relatively high number of GIT bacterial genomes that have been sequenced; about 28% of all fully sequenced bacterial genomes are part of the Human Microbiome Project [196]. Nevertheless, the high percentage of CBM74 domains found in enzymes from GIT related bacteria may indicate that CBM74 fulfills a specific role in starch digestion in the intestinal tract. In the human GIT, most of the (soluble) starch from food is degraded by α -amylases and glucoamylases of the host organism. However, Resistant Starch (RS) is harder to degrade due to its crystallinity or due to complex formation, either occurring naturally or after food processing [8,12]. Under normal conditions RS is fermented completely by microorganisms in the colon of the host [8,12].

Resistant Starch can be divided into five different types (RS1-5). The higher the RS number, the lower the degradation rate by human α -amylases in the GIT. RS3, also known as retrograded starch, is of special interest since it is formed without any additions and resists regular food processing, or is even formed during processing [12,197]. It is well known that SBDs greatly enhance the ability of α -amylases to degrade granular starches [154]. Since most CBM74 homologs are found in large α -amylases with additional SBDs it appears likely that CBM74

plays a role in resistant (granular) starch binding.

The ability of *Bifidobacteria* to degrade RS has been demonstrated in literature. Animal studies in which rats colonized with human microflora were fed a high RS diet showed that the number of *Bifidobacteria* and *Lactobacilli* in the microflora increased 10 to 100 fold when compared to a high sucrose diet, demonstrating a link between RS fermentation and representation of these two genera [198]. As shown in Figure 5, CBM74 is present in α -amylases from 22 different *Bifidobacterium* strains, constituting over 45% of all sequenced *Bifidobacterium* strains listed in GenBank (November 2015). Another study showed that *Ruminococcus bromii* L2-63 and *Bifidobacterium adolescentis* L2-32 individually are able to degrade especially RS3 up to about 50%, and even up to > 90% when co-cultured. In an obese test subject with a low percentage of RS fermentation, both *R. bromii* and *B. adolescentis* were absent from the microflora [126]. Addition of *B. adolescentis* L2-32 or *R. bromii* L2-63 improved RS3 fermentation with ~20% and ~45%, respectively, the latter restoring fermentation to levels similar to those of healthy volunteers. Proteins containing CBM74 homologs are present in the genomes of both these strains (WP_015523730.1 and EDN82501.1 in Fig. 5), but absent in the genomes of two other strains used in the same study which were unable to improve RS3 degradation significantly (*Eubacterium rectale* A1-86 and *Bacteroides thetaiotaomicron* 5482) [126].

The relative abundance of CBM74 in mammalian gut *Bifidobacterium* α -amylases is taken to suggest that CBM74 has a major role in degradation of RS in the mammalian GIT. The presence and proper functioning of this CBM74 domain thus may have strong effects on the efficiency of mammalian food digestion.

CBM74 thus may assist MaAmyA in the degradation of RS through binding to it. The binding of CBM74 to starch granules has been demonstrated experimentally (Fig. 3). In addition to binding, CBM74 may also be involved in preparation of the substrate (granule) surface for degradation, in a similar way as it is seen for CBMs in cellulase enzymes, where CBMs assist in unwinding the carbohydrate chains, making them more accessible for the action of the catalytic domain [47]. As shown, the presence of CBM74 results in formation of larger pores in starch granules (Fig. 1).

Our bioinformatics analysis revealed 77 CBM74 homologs in databases and confirmed that CBM74 constitutes a single domain. The CBM74 homologs clustered together in a phylogenetic analysis (Fig. 5) and showed low identity to other known CBMs. We therefore conclude that CBM74 represents a novel starch binding CBM family.

Acknowledgements

This study was partly funded by the Top Institute of Food & Nutrition (project B1003) and by the University of Groningen. ALvB is funded by an NWO Veni Grant .

